Take Home Project for Enterprise Business Analytics Module

By

SSEMWEZI BENJAMIN

REG No: 2023-M132-21506

TASK 1

Project Objectives

1. Compare key indicators across different regions and income groups to identify disparities and common trends.
2. Examine birth and death rates across different regions and countries to understand demographic shifts and population dynamics.
3. Analyze population density data to explore its relationship with other economic and social indicators.
4. Investigate trends in infant mortality rates and life expectancy at birth to assess public health improvements or challenges.
5. Analyze unemployment rates across regions and countries to understand labor market conditions and trends.
6. Analyze the relationship between electric power consumption and economic indicators such as GDP and GDP per capita.
7. Perform year-over-year comparisons to highlight significant changes in the dataset.
8. Provide recommendations for policy improvements based on data-driven insights.

Data Sources and Collection Methods

The dataset originates from the World Bank, an international financial institution providing financial and technical assistance to developing countries.

Dataset Representation

This dataset represents a comprehensive collection of key socio-economic indicators from countries worldwide. The data spans multiple years and includes metrics that can help analyze the economic, health, and social conditions of each country. The indicators can be used to understand regional disparities, economic development, public health, and infrastructure status. The dataset contains socio-economic indicators collected from 135 countries across the world from 1991 to 2008. It includes the following columns:

1. **Country Name**: The name of the country.
2. **Country Code**: The ISO code of the country.
3. **Region**: The region in which the country is located (e.g., Europe & Central Asia, Sub-Saharan Africa, Latin America and the Carrabean, Middle East and North Africa, East Asia and the Pacific, South Asia and North America).
4. **IncomeGroup**: The income classification of the country (e.g., High income: OECD, High Income(nonOECD),Upper middle income, Low income and lower middle income).
5. **Year**: The year the data was collected. It includes years from 1991 to 2014.
6. **Birth rate, crude (per 1,000 people)**: The number of live births per 1,000 people in the population.
7. **Death rate, crude (per 1,000 people)**: The number of deaths per 1,000 people in the population.
8. **Electric power consumption (kWh per capita)**: The average kilowatt-hour (kWh) of electric power consumed per person.
9. **GDP (USD)**: The Gross Domestic Product in U.S. dollars.
10. **GDP per capita (USD)**: The GDP divided by the population, indicating the average economic output per person.
11. **Individuals using the Internet (% of population)**: The percentage of the population that uses the internet.
12. **Infant mortality rate (per 1,000 live births)**: The number of infants dying before reaching one year of age, per 1,000 live births in a given year.
13. **Life expectancy at birth (years)**: The average number of years a newborn is expected to live, assuming mortality patterns at the time of birth remain constant.
14. **Population density (people per sq. km of land area)**: The number of people living per square kilometer of land area.
15. **Unemployment (% of total labor force)**: The percentage of the total labor force that is unemployed but actively seeking employment and willing to work.

The data collection methods include:

1. **Government Statistics:** Collected from national statistics offices and other relevant government agencies.
2. **International Organizations:** Compiled by international organizations such as the International Labour Organization (ILO), United Nations (UN), and World Health Organization (WHO).
3. **Surveys and Censuses:** Data obtained through national surveys and population censuses.
4. **Estimates and Projections:** Calculated using statistical models and estimation techniques when direct data collection is not feasible.

Relevance to Business Analytics

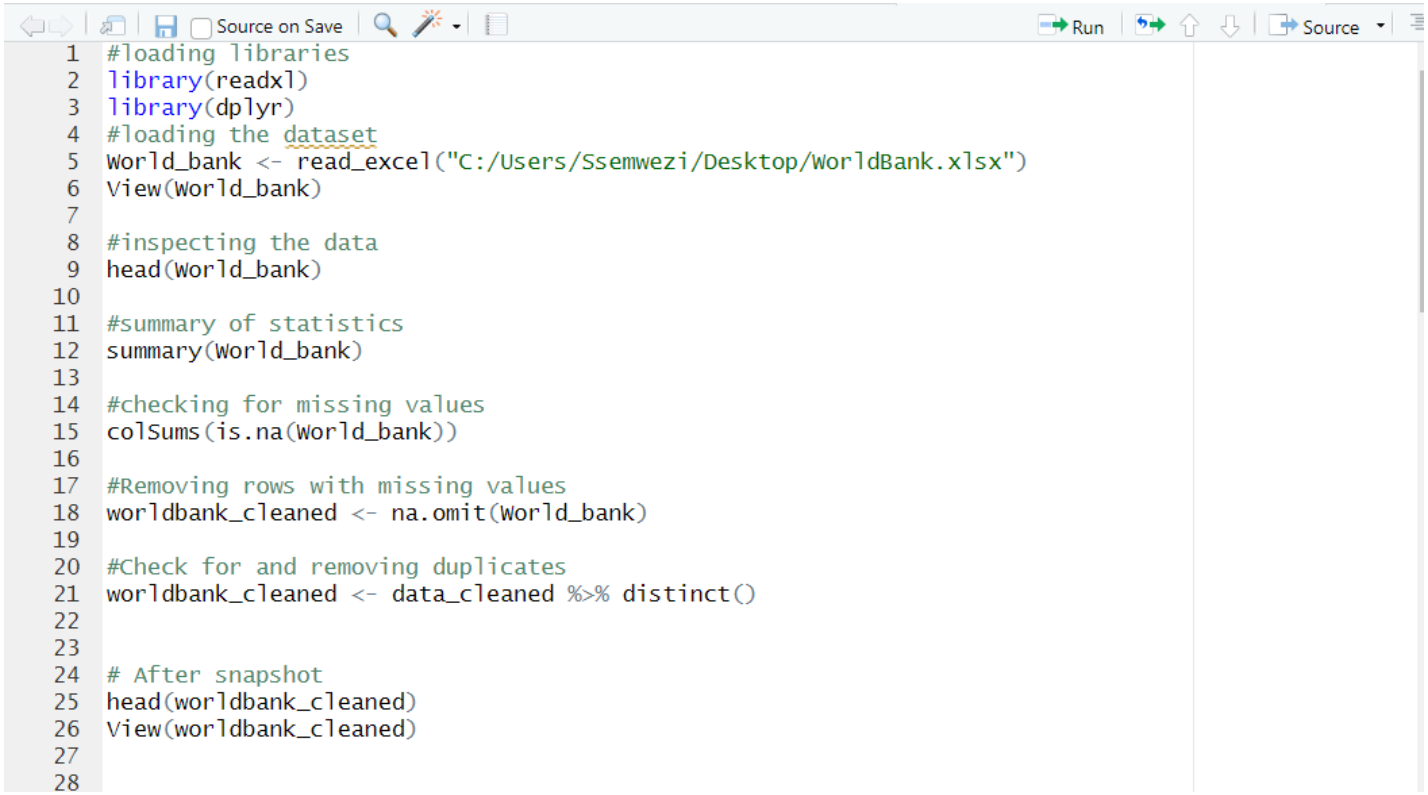
This dataset is relevant to business analytics for several reasons:

- **Economic Analysis**: Assessing GDP and GDP per capita to understand economic performance.
- **Health Studies:** Analyzing infant mortality rate, life expectancy, and crude birth/death rates to gauge public health.
- **Infrastructure Planning**: Using electric power consumption and internet usage data to plan and develop infrastructure projects.
- **Social Programs**: Identifying regions with high unemployment rates to target job creation programs.
- **Demographic Studies**: Studying population density and birth rates for urban planning and resource allocation.

TASK 2:

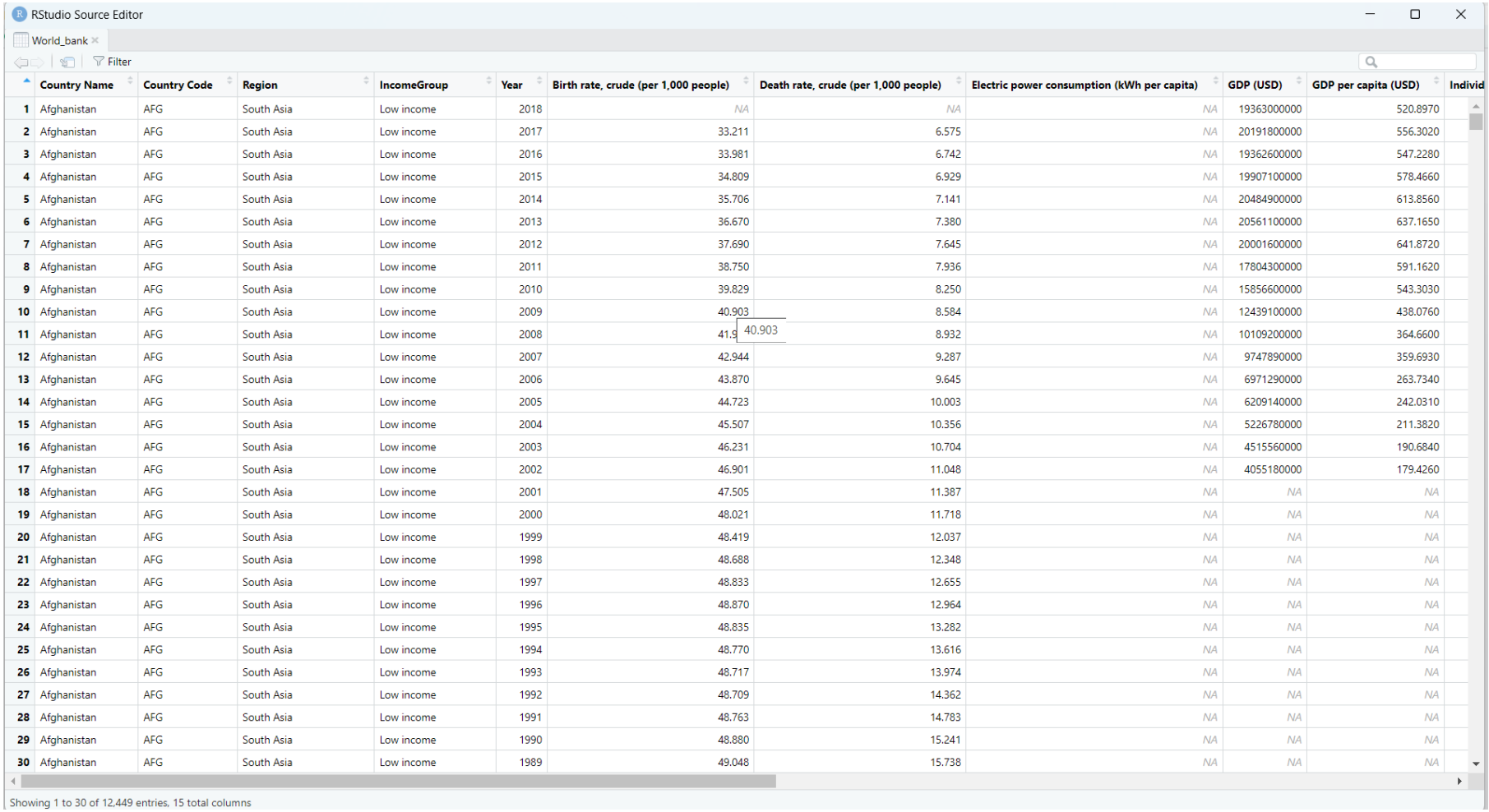
Data Cleaning Using R

Data cleaning on the dataset was done in R, it included handling missing values, removing duplicates, correcting errors, and standardizing data formats. Below are the steps taken in R and before and after snapshots of the dataset.



```r
#loading libraries
library(readxl)
library(dplyr)
#loading the dataset
World_bank <- read_excel("C:/Users/Ssemwezi/Desktop/WorldBank.xlsx")
View(World_bank)

#inspecting the data
head(World_bank)

#summary of statistics
summary(World_bank)

#checking for missing values
colSums(is.na(World_bank))

#Removing rows with missing values
worldbank_cleaned <- na.omit(World_bank)

#Check for and removing duplicates
worldbank_cleaned <- data_cleaned %>% distinct()


# After snapshot
head(worldbank_cleaned)
View(worldbank_cleaned)

```

Dataset before cleaning


RStudio Source Editor

World_bank

Filter

	Country Name	Country Code	Region	IncomeGroup	Year	Birth rate, crude (per 1,000 people)	Death rate, crude (per 1,000 people)	Electric power consumption (kWh per capita)	GDP (USD)	GDP per capita (USD)	Individ
1	Afghanistan	AFG	South Asia	Low income	2018	NA	NA	NA	19363000000	520.8970	
2	Afghanistan	AFG	South Asia	Low income	2017	33.211	6.575	NA	20191800000	556.3020	
3	Afghanistan	AFG	South Asia	Low income	2016	33.981	6.742	NA	19362600000	547.2280	
4	Afghanistan	AFG	South Asia	Low income	2015	34.809	6.929	NA	19907100000	578.4660	
5	Afghanistan	AFG	South Asia	Low income	2014	35.706	7.141	NA	20484900000	613.8560	
6	Afghanistan	AFG	South Asia	Low income	2013	36.670	7.380	NA	20561100000	637.1650	
7	Afghanistan	AFG	South Asia	Low income	2012	37.690	7.645	NA	20001600000	641.8720	
8	Afghanistan	AFG	South Asia	Low income	2011	38.750	7.936	NA	17804300000	591.1620	
9	Afghanistan	AFG	South Asia	Low income	2010	39.829	8.250	NA	15856600000	543.3030	
10	Afghanistan	AFG	South Asia	Low income	2009	40.903	8.584	NA	12439100000	438.0760	
11	Afghanistan	AFG	South Asia	Low income	2008	41.9	8.932	NA	10109200000	364.6600	
12	Afghanistan	AFG	South Asia	Low income	2007	42.944	9.287	NA	9747890000	359.6930	
13	Afghanistan	AFG	South Asia	Low income	2006	43.870	9.645	NA	6971290000	263.7340	
14	Afghanistan	AFG	South Asia	Low income	2005	44.723	10.003	NA	6209140000	242.0310	
15	Afghanistan	AFG	South Asia	Low income	2004	45.507	10.356	NA	5226780000	211.3820	
16	Afghanistan	AFG	South Asia	Low income	2003	46.231	10.704	NA	4515560000	190.6840	
17	Afghanistan	AFG	South Asia	Low income	2002	46.901	11.048	NA	4055180000	179.4260	
18	Afghanistan	AFG	South Asia	Low income	2001	47.505	11.387	NA	NA	NA	
19	Afghanistan	AFG	South Asia	Low income	2000	48.021	11.718	NA	NA	NA	
20	Afghanistan	AFG	South Asia	Low income	1999	48.419	12.037	NA	NA	NA	
21	Afghanistan	AFG	South Asia	Low income	1998	48.688	12.348	NA	NA	NA	
22	Afghanistan	AFG	South Asia	Low income	1997	48.833	12.655	NA	NA	NA	
23	Afghanistan	AFG	South Asia	Low income	1996	48.870	12.964	NA	NA	NA	
24	Afghanistan	AFG	South Asia	Low income	1995	48.835	13.282	NA	NA	NA	
25	Afghanistan	AFG	South Asia	Low income	1994	48.770	13.616	NA	NA	NA	
26	Afghanistan	AFG	South Asia	Low income	1993	48.717	13.974	NA	NA	NA	
27	Afghanistan	AFG	South Asia	Low income	1992	48.709	14.362	NA	NA	NA	
28	Afghanistan	AFG	South Asia	Low income	1991	48.763	14.783	NA	NA	NA	
29	Afghanistan	AFG	South Asia	Low income	1990	48.880	15.241	NA	NA	NA	
30	Afghanistan	AFG	South Asia	Low income	1989	49.048	15.738	NA	NA	NA	

40.903

Showing 1 to 30 of 12,449 entries, 15 total columns

Dataset had 12,449 entries before cleaning, these reduced to 2,775 after cleaning and removing all the blank entries

Dataset after cleaning

RStudio Source Editor

worldbank_cleaned

Filter

	Country Name	Country Code	Region	IncomeGroup	Year	Birth rate, crude (per 1,000 people)	Death rate, crude (per 1,000 people)	Electric power consumption (kWh per capita)	GDP (USD)	GDP per capita (USD)	Indiv
1	Albania	ALB	Europe & Central Asia	Upper middle income	2014	12.259	7.219	2309.3700	1.32282e+10	4578.670	
2	Albania	ALB	Europe & Central Asia	Upper middle income	2013	12.257	7.096	2533.2500	1.27763e+10	4413.080	
3	Albania	ALB	Europe & Central Asia	Upper middle income	2012	12.197	6.996	2118.3300	1.23198e+10	4247.610	
4	Albania	ALB	Europe & Central Asia	Upper middle income	2011	12.100	6.915	2205.7000	1.28909e+10	4437.180	
5	Albania	ALB	Europe & Central Asia	Upper middle income	2010	12.001	6.841	1943.3400	1.19270e+10	4094.360	
6	Albania	ALB	Europe & Central Asia	Upper middle income	2009	11.945	6.756	1835.6800	1.20442e+10	4114.130	
7	Albania	ALB	Europe & Central Asia	Upper middle income	2008	11.973	6.650	1617.7400	1.28814e+10	4370.540	
8	Albania	ALB	Europe & Central Asia	Upper middle income	2007	12.118	6.518	1213.1200	1.06773e+10	3595.040	
9	Albania	ALB	Europe & Central Asia	Upper middle income	2006	12.398	6.365	1218.3600	8.89607e+09	2972.740	
10	Albania	ALB	Europe & Central Asia	Upper middle income	2005	12.821	6.206	1722.0700	8.05208e+09	2673.790	
11	Albania	ALB	Europe & Central Asia	Upper middle income	2004	13.381	6.061	1797.5300	7.18468e+09	2373.580	
12	Albania	ALB	Europe & Central Asia	Upper middle income	2003	14.048	5.952	1469.2600	5.61149e+09	1846.120	
13	Albania	ALB	Europe & Central Asia	Upper middle income	2002	14.790	5.891	1578.1700	4.34807e+09	1425.120	
14	Albania	ALB	Europe & Central Asia	Upper middle income	2001	15.590	5.879	1351.2300	3.92210e+09	1281.660	
15	Albania	ALB	Europe & Central Asia	Upper middle income	2000	16.436	5.914	1449.6500	3.48036e+09	1126.680	
16	Albania	ALB	Europe & Central Asia	Upper middle income	1999	17.321	5.983	1414.7000	3.21212e+09	1033.240	
17	Albania	ALB	Europe & Central Asia	Upper middle income	1998	18.238	6.067	734.8500	2.54597e+09	813.790	
18	Albania	ALB	Europe & Central Asia	Upper middle income	1997	19.173	6.147	694.6650	2.25852e+09	717.381	
19	Albania	ALB	Europe & Central Asia	Upper middle income	1996	20.106	6.213	904.3470	3.19964e+09	1009.980	
20	Albania	ALB	Europe & Central Asia	Upper middle income	1995	21.020	6.250	663.7840	2.39276e+09	750.604	
21	Algeria	DZA	Middle East & North Africa	Upper middle income	2014	25.538	4.709	1362.8700	2.14000e+11	5493.060	
22	Algeria	DZA	Middle East & North Africa	Upper middle income	2013	25.451	4.699	1278.9200	2.10000e+11	5499.590	
23	Algeria	DZA	Middle East & North Africa	Upper middle income	2012	25.246	4.685	1237.9700	2.09000e+11	5592.220	
24	Algeria	DZA	Middle East & North Africa	Upper middle income	2011	24.934	4.670	1123.3300	2.00000e+11	5455.840	
25	Algeria	DZA	Middle East & North Africa	Upper middle income	2010	24.504	4.656	1016.6400	1.61000e+11	4480.790	
26	Algeria	DZA	Middle East & North Africa	Upper middle income	2009	23.932	4.643	866.3070	1.37000e+11	3883.270	
27	Algeria	DZA	Middle East & North Africa	Upper middle income	2008	23.226	4.637	947.2910	1.71000e+11	4923.630	
28	Algeria	DZA	Middle East & North Africa	Upper middle income	2007	22.424	4.639	894.2850	1.35000e+11	3950.510	
29	Algeria	DZA	Middle East & North Africa	Upper middle income	2006	21.582	4.651	862.4300	1.17000e+11	3478.710	
30	Algeria	DZA	Middle East & North Africa	Upper middle income	2005	20.774	4.674	890.6260	1.03000e+11	3113.090	
31	Algeria	DZA	Middle East & North Africa	Upper middle income	2004	20.077	4.710	804.1070	8.53250e+10	2609.950	

Showing 1 to 32 of 2,775 entries, 15 total columns

TASK 3

Rationale for Integration

Integrating additional socio-economic indicators such as education and healthcare expenditure with the original dataset provides a more comprehensive view of a country's socio-economic status. This integration allows for deeper analysis and insights into how various factors like education and healthcare influence economic indicators such as GDP, unemployment, and life expectancy.

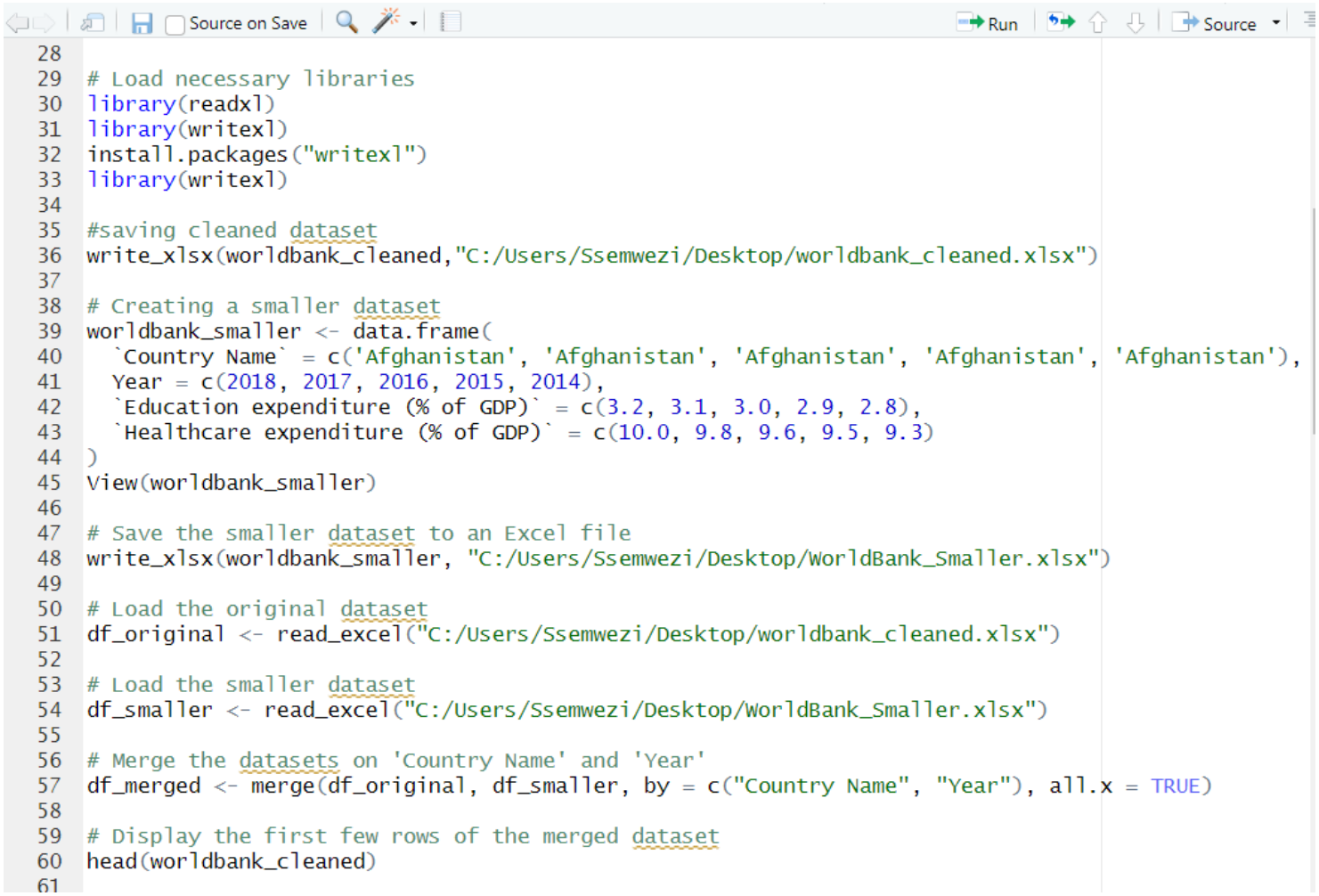
Process of Data Integration

1. **Data Preparation**: Created a smaller dataset with relevant additional indicators.
2. **Loading Datasets**: Loaded both the original and smaller datasets into R.
3. **Merging Datasets**: Merged the datasets using common keys (Country Name and Year).

Challenges Faced

1. **Data Alignment**: Ensuring that the keys used for merging (country names and years) are consistent across both datasets.
2. **Handling Missing Values**: The smaller dataset might not have data for all countries and years present in the original dataset, resulting in missing values after the merge.
3. **Data Consistency**: Ensuring that the data formats and types are consistent across datasets to avoid integration issues.

R code for the entire process

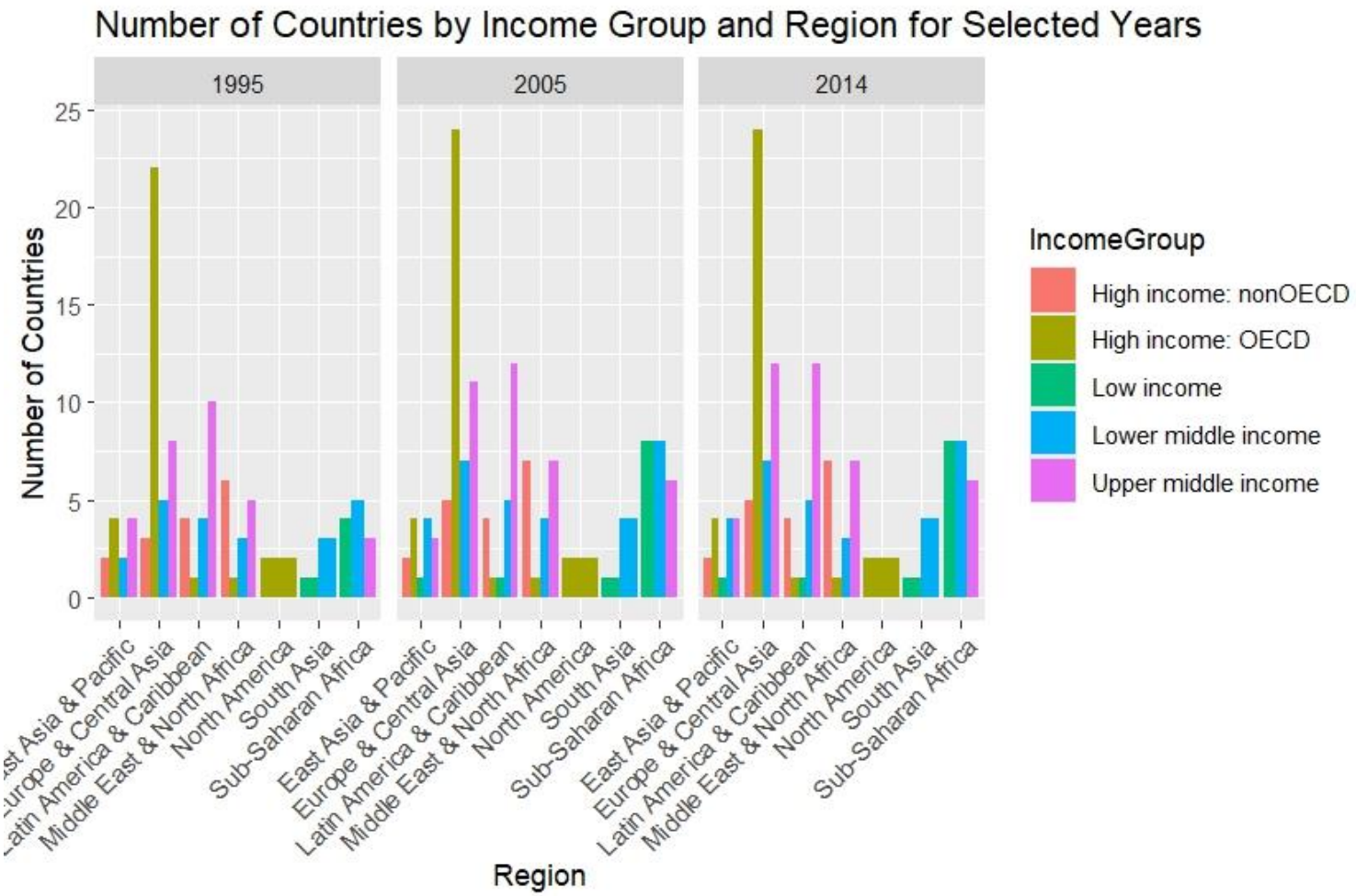


```r

# Load necessary libraries
library(readxl)
library(writexl)
install.packages("writexl")
library(writexl)

#saving cleaned dataset
write_xlsx(worldbank_cleaned,"C:/Users/Ssemwezi/Desktop/worldbank_cleaned.xlsx")

# Creating a smaller dataset
worldbank_smaller <- data.frame(
  `Country Name` = c('Afghanistan', 'Afghanistan', 'Afghanistan', 'Afghanistan', 'Afghanistan'),
  Year = c(2018, 2017, 2016, 2015, 2014),
  `Education expenditure (% of GDP)` = c(3.2, 3.1, 3.0, 2.9, 2.8),
  `Healthcare expenditure (% of GDP)` = c(10.0, 9.8, 9.6, 9.5, 9.3)
)
View(worldbank_smaller)

# Save the smaller dataset to an Excel file
write_xlsx(worldbank_smaller, "C:/Users/Ssemwezi/Desktop/WorldBank_Smaller.xlsx")

# Load the original dataset
df_original <- read_excel("C:/Users/Ssemwezi/Desktop/worldbank_cleaned.xlsx")

# Load the smaller dataset
df_smaller <- read_excel("C:/Users/Ssemwezi/Desktop/WorldBank_Smaller.xlsx")

# Merge the datasets on 'Country Name' and 'Year'
df_merged <- merge(df_original, df_smaller, by = c("Country Name", "Year"), all.x = TRUE)

# Display the first few rows of the merged dataset
head(worldbank_cleaned)

```

TASK 4

Summary of statistics and data visualization

1. Barplot representing number of countries by income group and region



Interpretation of the Plot:

The plot visualizes the number of countries per income group across various regions for the years 1995, 2005, and 2014. The income groups include High income (nonOECD), High income (OECD), Low income, Lower middle income, and Upper middle income. The plot is faceted by year, showing separate bar charts for each of the selected years.

Detailed Breakdown

1. **East Asia & Pacific**:
    - **1995**: Predominantly lower middle income with a few upper middle income and high income (nonOECD) countries.
    - **2005**: Increase in upper middle-income countries and a slight decrease in lower middle-income countries.
    - **2014**: Further increase in upper middle-income countries, maintaining a stable number of lower middle-income countries.
2. **Europe & Central Asia**:
    - **1995**: Majority are high income (nonOECD) with a few lower middle-income and upper middle-income countries.
    - **2005**: Noticeable increase in upper middle-income countries.
    - **2014**: Further increase in upper middle-income countries and a stable number of high income (nonOECD) countries.
3. **Latin America & Caribbean**:
    - **1995**: A balanced mix of lower middle income and upper middle-income countries, with some high income (nonOECD) countries.
    - **2005**: Increase in upper middle-income countries.
    - **2014**: Continued increase in upper middle-income countries and a slight decrease in lower middle-income countries.
4. **Middle East & North Africa**:
    - **1995**: Predominantly upper middle-income countries with a few lower middle-income countries.
    - **2005**: Increase in upper middle-income countries and stable lower middle-income countries.
    - **2014**: Continued dominance of upper middle-income countries.
5. **North America**:
    - **1995, 2005, 2014**: Consistently high income (OECD) with no significant change over the years.
6. **South Asia**:
    - **1995**: Primarily low income and lower middle-income countries.
    - **2005**: Increase in lower middle-income countries and a decrease in low income countries.

- **2014**: Further decrease in low income countries, with a significant increase in lower middle-income countries.

7. **Sub-Saharan Africa**:
   - **1995**: Predominantly low income with a few lower middle-income countries.
   - **2005**: Stable number of low-income countries and a slight increase in lower middle-income countries.
   - **2014**: Continued dominance of low-income countries with a noticeable increase in lower middle-income countries.

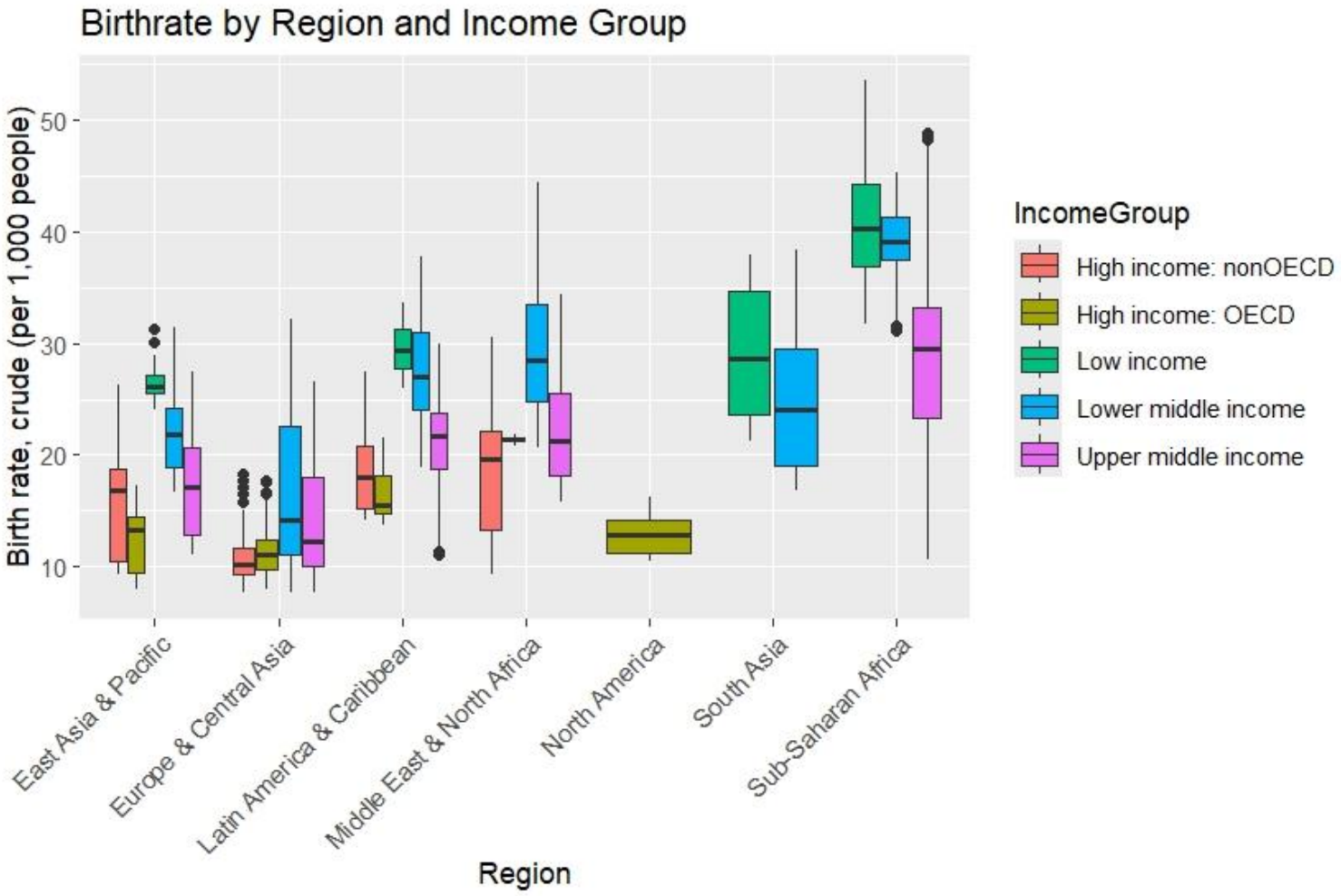
General Observations

1. **Income Group Distribution Over Time**:
   - Across all regions, there is a noticeable trend of countries transitioning from low income to lower middle income and from lower middle income to upper middle income over the years.
   - High income (OECD and nonOECD) countries remain relatively stable, indicating that countries in these groups have maintained their economic status.
2. **Regional Economic Development**:
   - **East Asia & Pacific** and **Europe & Central Asia** show significant economic development, with an increasing number of countries moving to higher income groups over the years.
   - **Sub-Saharan Africa** remains predominantly low income, although there is a slight increase in lower middle income countries, indicating gradual economic development.
   - **South Asia** shows a similar trend of economic improvement, with a shift from low income to lower middle income.
3. **Stability in High-Income Regions**:
   - Regions like **North America** and parts of **Europe & Central Asia** show stability in their high-income status, reflecting sustained economic prosperity.

Conclusion

The plot highlights the economic progression of various regions over the selected years, showcasing a general trend of upward mobility in income groups for many countries. However, it also points out the persistent economic challenges faced by regions like Sub-Saharan Africa, which remain predominantly low income. This visualization underscores the need for region-specific economic policies and interventions to promote balanced global economic development.

2. Boxplot showing Birthrate by Region and income Group



Interpretation of the Chart:
The chart visualizes birth rates (crude, per 1,000 people) across different regions and income groups. Each box plot represents the distribution of birth rates for a particular region and income group.

Detailed Breakdown by Region and Income Group

1. **East Asia & Pacific**:
    - **High income (nonOECD)**: Lower birth rates, with a small range.
    - **Lower middle income**: Higher birth rates compared to high income.
    - **Upper middle income**: Moderate birth rates, with some variability.
2. **Europe & Central Asia**:
    - **High income (nonOECD)**: Lower birth rates, consistent with high income status.
    - **Lower middle income**: Higher birth rates than high income, but relatively low compared to other regions.
    - **Upper middle income**: Moderate birth rates, higher than high income but lower than lower middle income.
3. **Latin America & Caribbean**:
    - **High income (nonOECD)**: Lower birth rates.
    - **Lower middle income**: Moderate birth rates.
    - **Upper middle income**: Higher birth rates compared to other income groups in this region.
4. **Middle East & North Africa**:
    - **Upper middle income**: High birth rates, indicating a younger population on average.
5. **North America**:
    - **High income (OECD)**: Lower birth rates, consistent with developed country trends.
6. **South Asia**:
    - **Low income**: High birth rates, indicating higher fertility rates.
    - **Lower middle income**: Slightly lower but still high birth rates.
    - **Upper middle income**: Moderate birth rates, showing economic improvement and associated changes in birth rates.

7. **Sub-Saharan Africa**:
    - **Low income**: Very high birth rates, reflecting higher fertility rates in the region.
    - **Lower middle income**: Slightly lower but still high birth rates.
    - **Upper middle income**: Moderate birth rates, showing economic progress but still higher than other regions.

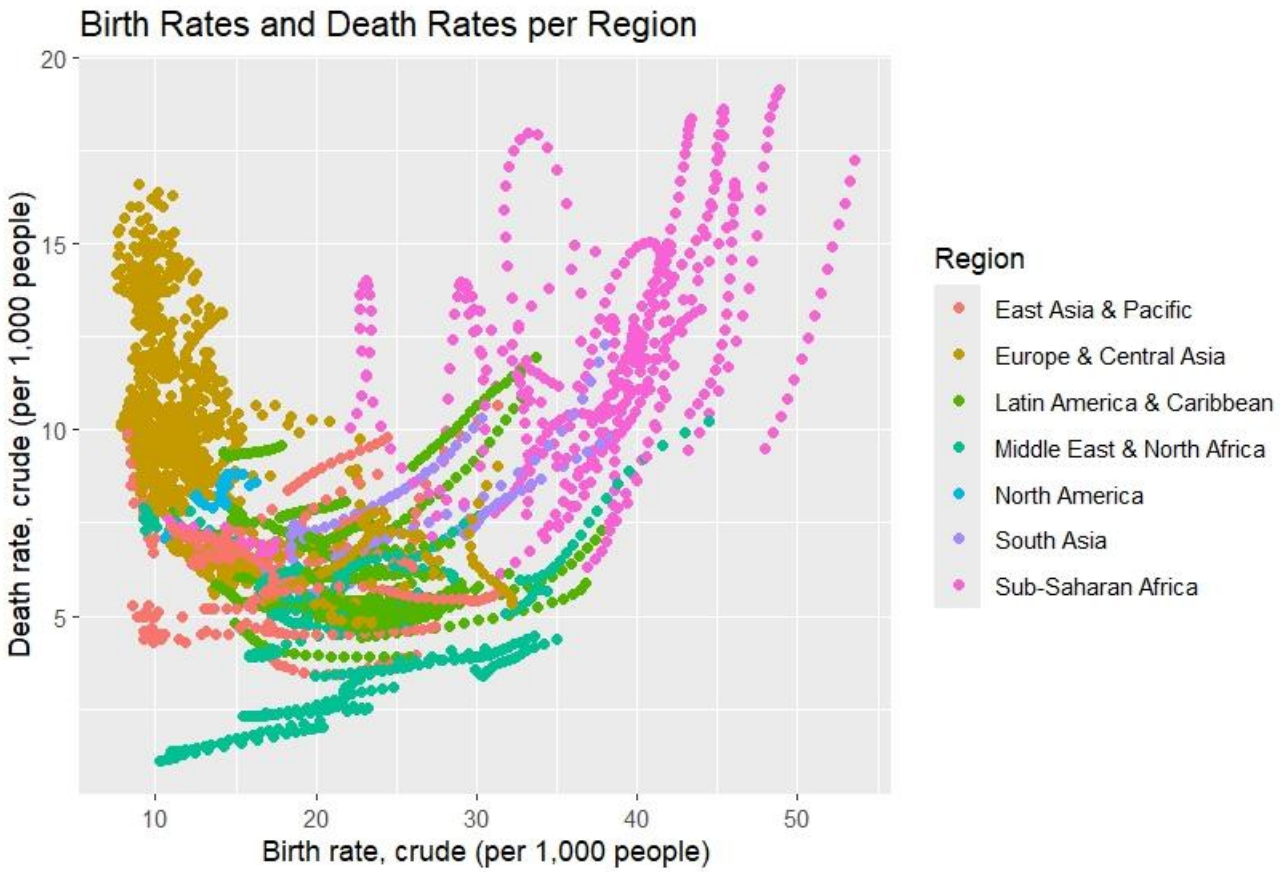
General Observations

1. **Income Groups**:
    - High income groups (both nonOECD and OECD) consistently show lower birth rates across all regions, reflecting lower fertility rates typical of more developed economies.
    - Low income groups have the highest birth rates, especially in regions like Sub-Saharan Africa and South Asia.
    - As income increases from low to lower middle and upper middle, birth rates tend to decrease, reflecting the demographic transition theory.
2. **Regional Differences**:
    - Sub-Saharan Africa and South Asia show the highest birth rates, particularly in low and lower middle income groups.
    - Regions like North America and parts of Europe & Central Asia show lower birth rates across all income groups, consistent with trends in more developed regions.
    - Middle East & North Africa, and Latin America & Caribbean show significant variability in birth rates across different income groups, indicating diverse demographic patterns within these regions.

Conclusion

The chart highlights the relationship between economic status (income group) and birth rates across different regions. It shows a clear trend where higher income is associated with lower birth rates, and lower income with higher birth rates. This trend is consistent across most regions, with Sub-Saharan Africa and South Asia showing the highest birth rates overall. The chart underscores the impact of economic development on demographic patterns and can be useful for policy-making and targeted interventions.

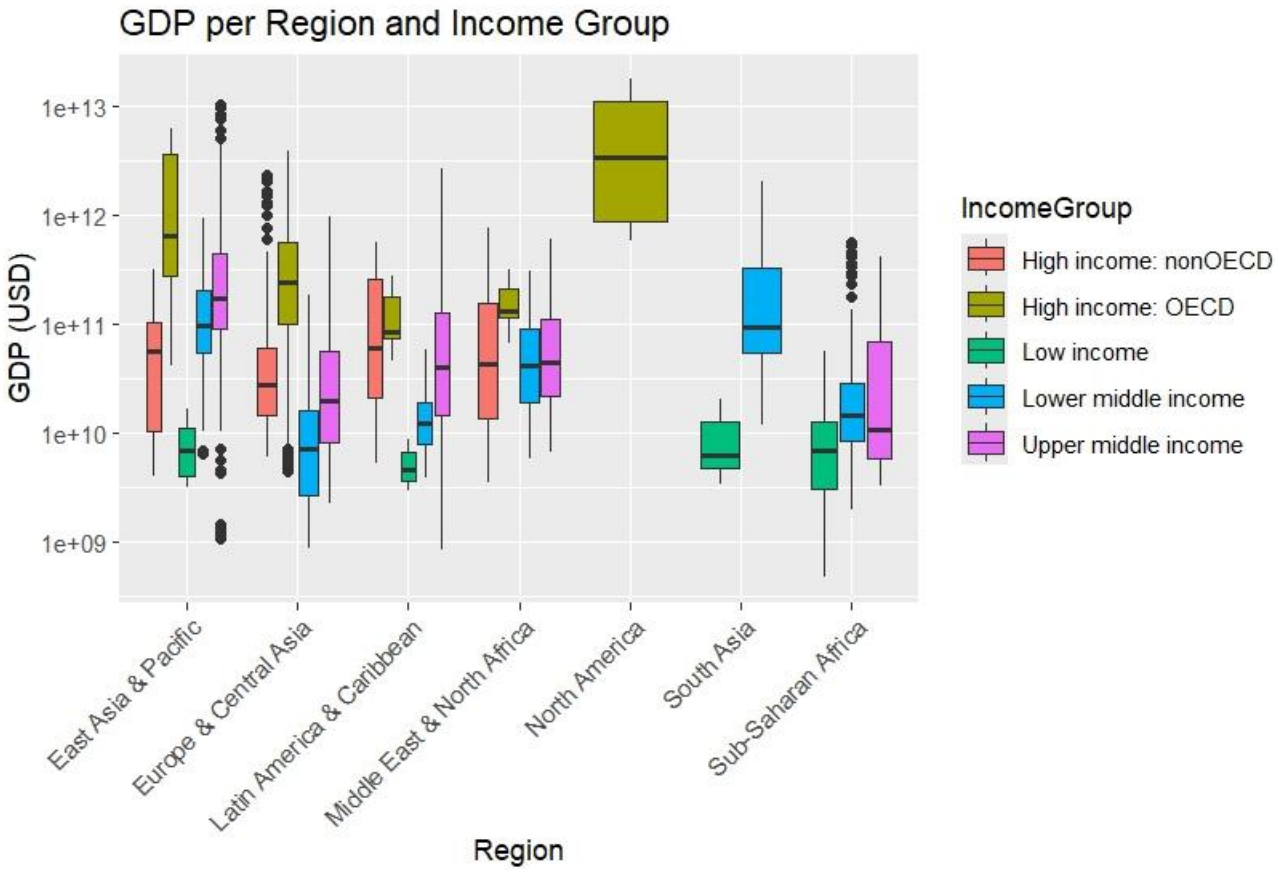
3. **Scatter plot showing birth rates and Death rates per Region**



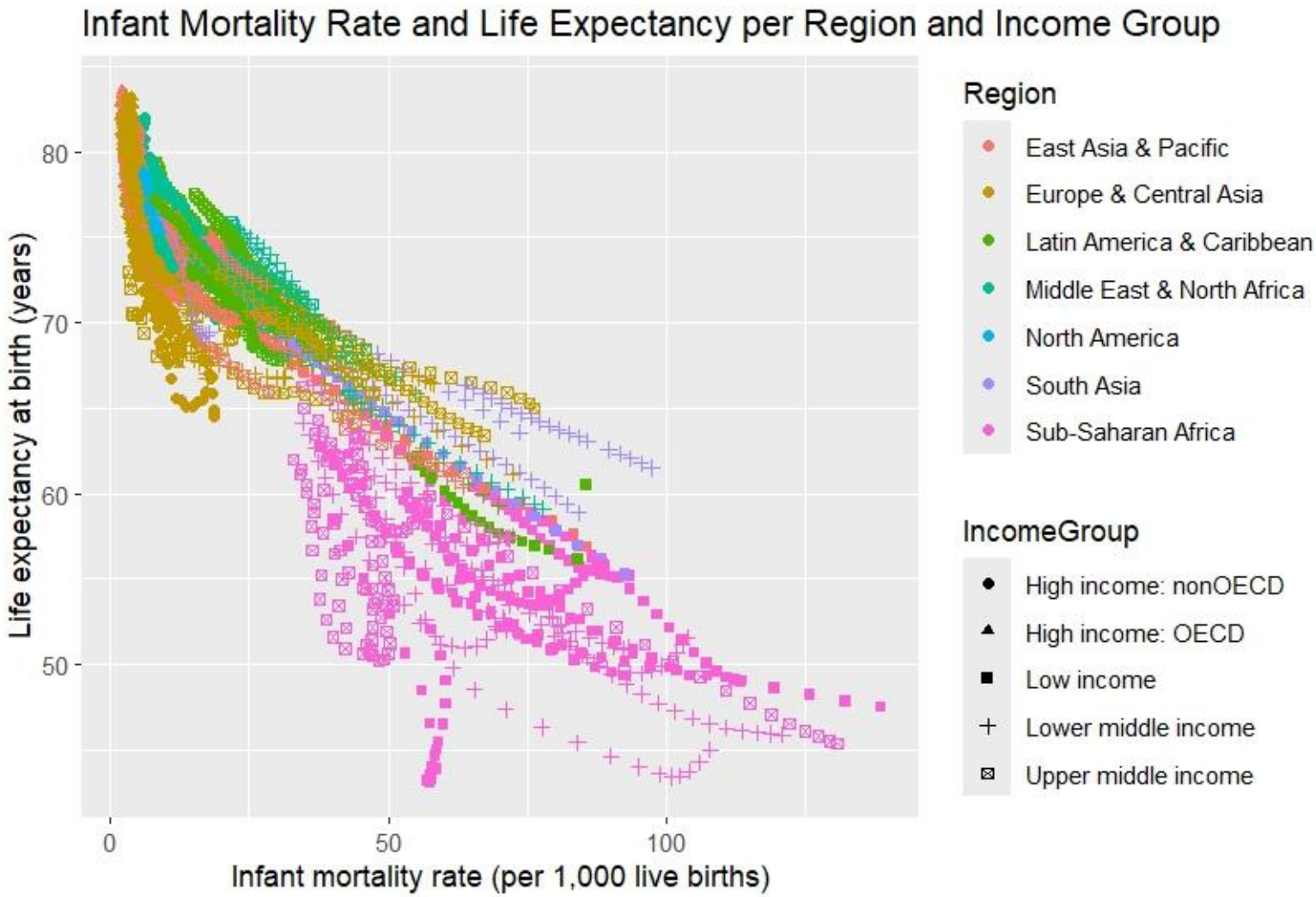
1. **GDP per Region and Income Group**:
    - **Interpretation**: The box plots show a wide range in GDP across different regions and income groups, with higher GDP in high-income regions and lower GDP in low-income regions.

- **Conclusion**: Economic disparity is evident across regions and income groups, indicating the need for economic policies aimed at promoting growth in lower-income regions.

4. Boxplot showing variations in GDP per region and income Group

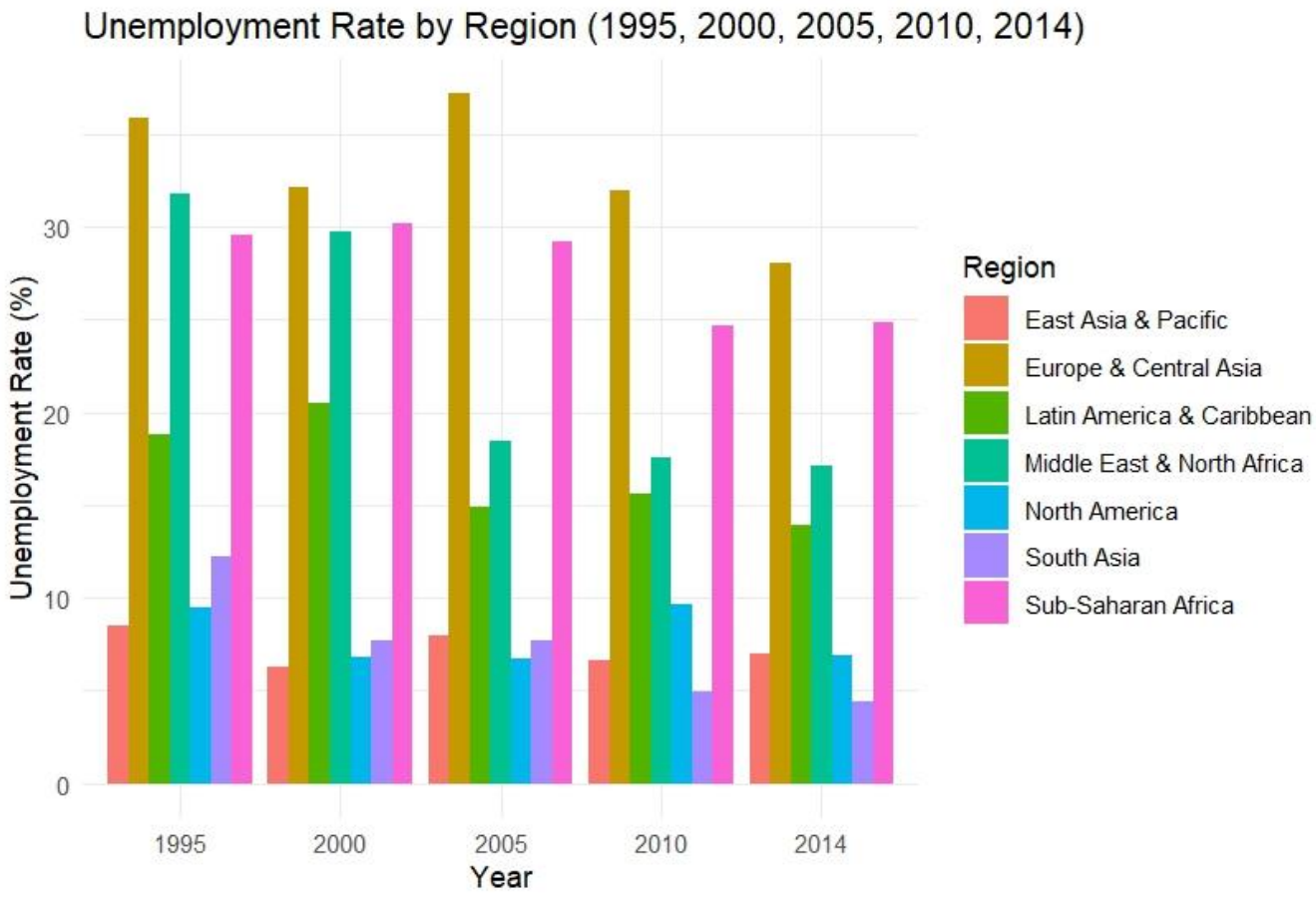


5. **Scatter plot showing variations in Infant Mortality Rate and Life Expectancy by Region and Income Group**:



- **Interpretation**: The scatter plot illustrates that regions with higher infant mortality rates generally have lower life expectancies. Sub-Saharan Africa has high infant mortality and low life expectancy, while high-income regions show the opposite trend.
- **Conclusion**: Improving infant mortality rates can significantly impact life expectancy, suggesting that healthcare improvements for infants and mothers are crucial for overall population health.

6. **Bar chart showing Unemployment Rate per Region per Year**:



This bar chart illustrates the unemployment rates across different regions for the years 1995, 2000, 2005, 2010, and 2014. Here's a detailed interpretation and analysis:

Key Observations by Region

1. **East Asia & Pacific (Brown)**:
    - Consistently low unemployment rates across all years, with a slight increase around 2000 and 2005.
2. **Europe & Central Asia (Orange)**:
    - Shows moderate unemployment rates, with a noticeable peak around 2005, suggesting economic challenges or transitions during that period.
3. **Latin America & Caribbean (Green)**:
    - Unemployment rates were relatively high in 1995 and 2000, followed by a decrease and then a slight increase in 2010 and 2014.
4. **Middle East & North Africa (Light Green)**:
    - This region displays a significant increase in unemployment rates in 2005 and 2010, indicating possible economic or social disruptions.
5. **North America (Blue)**:
    - Unemployment rates remained relatively low and stable, with a slight increase in 2010, possibly due to the global financial crisis of 2008-2009.
6. **South Asia (Pink)**:
    - Relatively low unemployment rates throughout the years, with a slight increase in 2010 and 2014.
7. **Sub-Saharan Africa (Magenta)**:
    - Exhibits the highest unemployment rates among all regions, with a peak in 2005 and consistent high levels throughout the years.

Analysis and Insights

1. **Economic Stability and Development**:
    - Regions like East Asia & Pacific and North America show relatively low unemployment rates, indicating stable economic conditions and effective labor market policies.

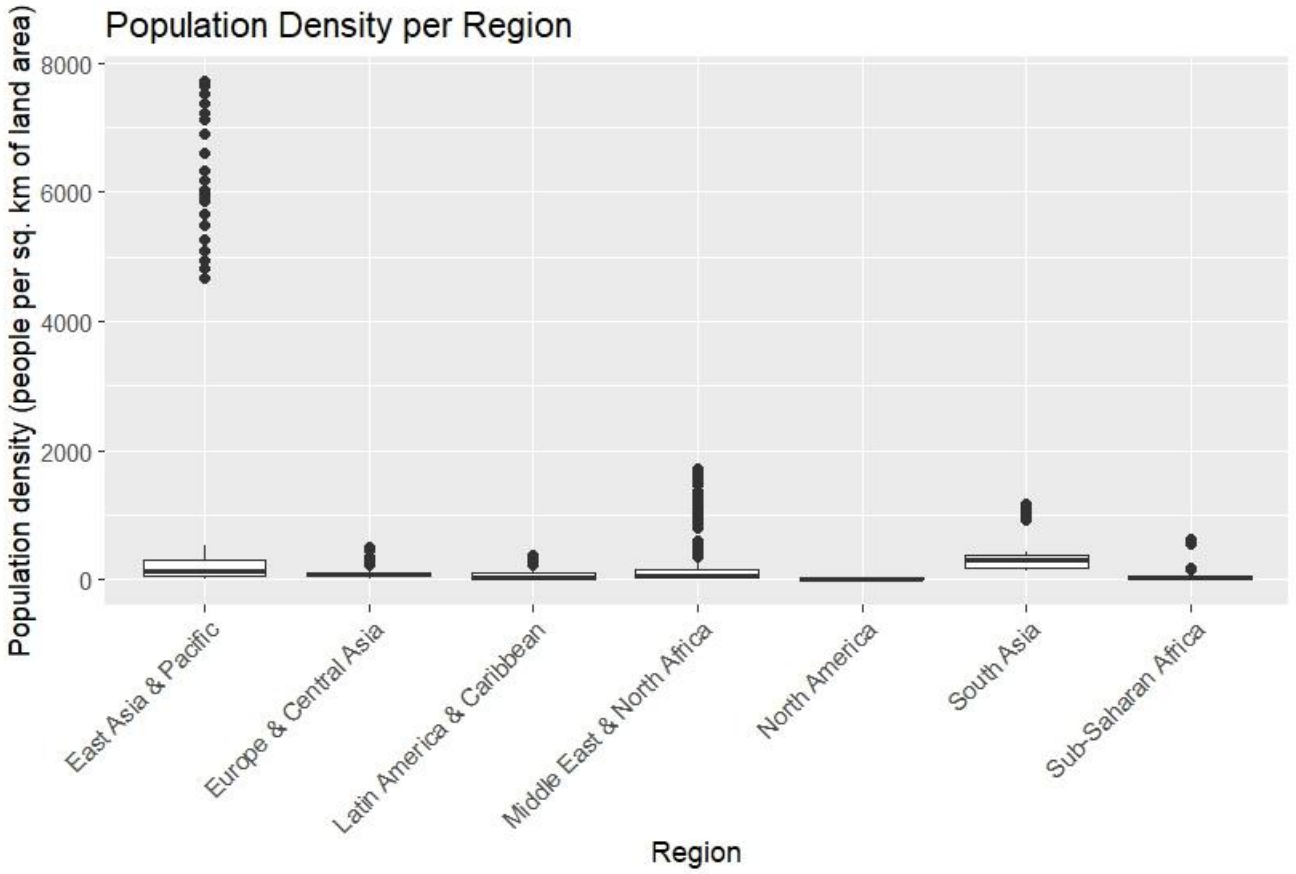
    - In contrast, Sub-Saharan Africa has consistently high unemployment rates, highlighting persistent economic challenges and possibly high youth unemployment.
2. **Impact of Global Events**:
    - The spike in unemployment in North America around 2010 can be linked to the aftermath of the global financial crisis.
    - Similar trends in Europe & Central Asia and Latin America & Caribbean suggest these regions were also affected by global economic downturns.
3. **Regional Specific Challenges**:
    - The Middle East & North Africa region shows significant variability in unemployment rates, which may be attributed to political instability, economic transitions, or social unrest during certain periods.
4. **Trends Over Time**:
    - Some regions like South Asia and East Asia & Pacific have shown resilience with relatively stable unemployment rates, while others like Sub-Saharan Africa continue to struggle with high unemployment.

7. **Population Density per Region**:
   The box plots highlight the variation in population density across regions, with regions like East Asia & Pacific showing higher population densities compared to others.

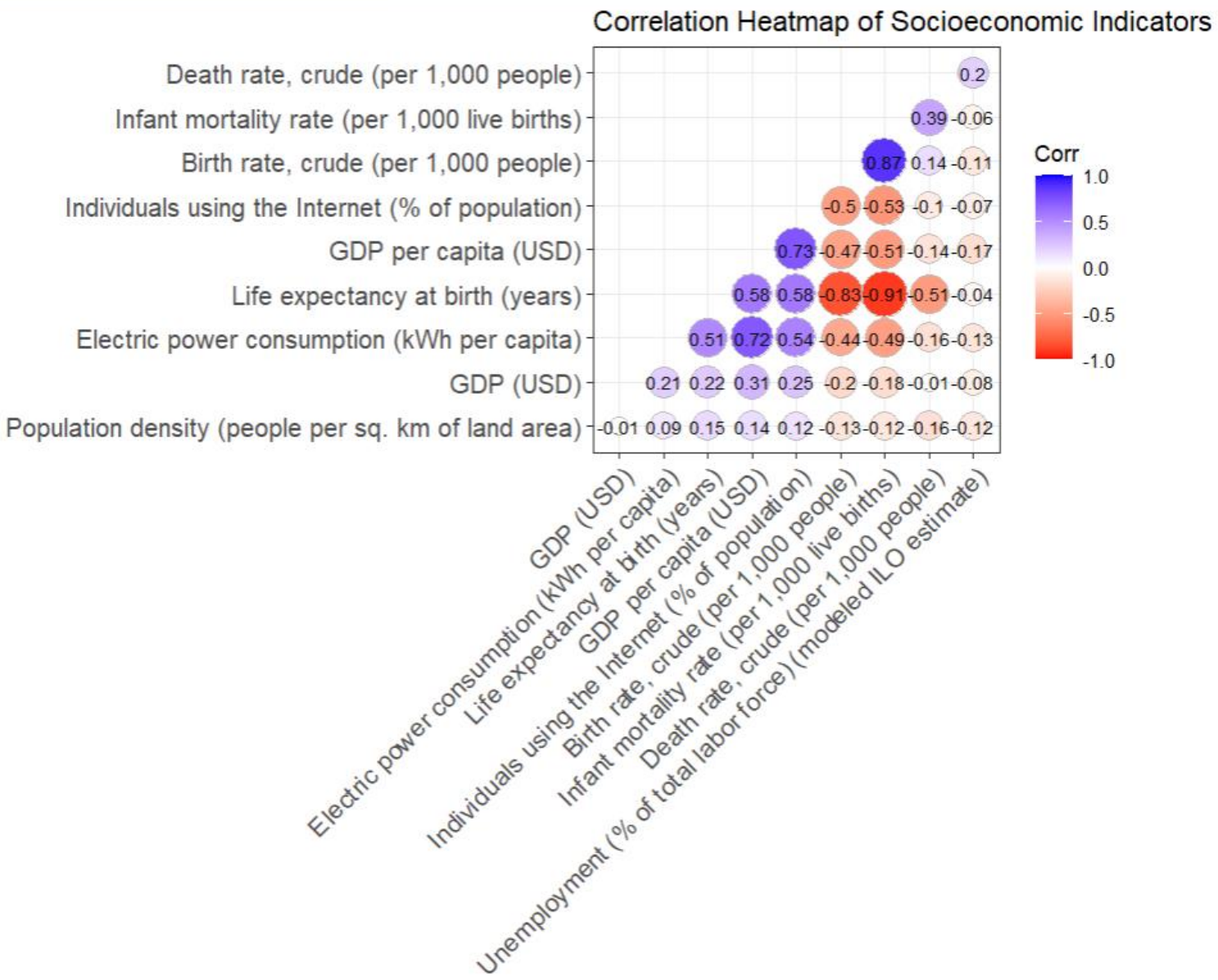
   Analysis:

- **High Population Density in Asia**: South Asia and East Asia & Pacific have high population densities due to large populations in relatively small land areas. This can be attributed to countries like India, Bangladesh, and China.
- **Urbanization and Population Pressure**: The presence of outliers in East Asia & Pacific indicates highly urbanized areas with extremely high population densities, possibly due to mega-cities or urban regions.
- **Comparative Density**: Other regions like Sub-Saharan Africa, North America, and Europe & Central Asia have lower and more consistent population densities, suggesting less population pressure on land in these regions.

Conclusion: Population density impacts infrastructure, resources, and urban planning. High-density regions may face more pressure on resources and services.



8. **Correlation Heatmap**: The heatmap provides a clear visualization of how socioeconomic indicators are interrelated. Strong positive correlations suggest areas where improvement in one indicator is likely to coincide with improvements in another. Negative correlations highlight potential areas of concern where increases in one indicator could lead to decreases in another. Understanding these relationships can help policymakers prioritize interventions that yield broad socioeconomic benefits.


Correlation Heatmap of Socioeconomic Indicators
Death rate, crude (per 1,000 people)
Infant mortality rate (per 1,000 live births)
Birth rate, crude (per 1,000 people)
Individuals using the Internet (% of population)
GDP per capita (USD)
Life expectancy at birth (years)
Electric power consumption (kWh per capita)
GDP (USD)
Population density (people per sq. km of land area)
0.2
0.39 -0.06
0.87 0.14 -0.11
-0.5 -0.53 -0.1 -0.07
0.73 -0.47 -0.51 -0.14 -0.17
0.58 0.58 -0.83 -0.91 -0.51 -0.04
0.51 0.72 0.54 -0.44 -0.49 -0.16 -0.13
0.21 0.22 0.31 0.25 -0.2 -0.18 -0.01 -0.08
-0.01 0.09 0.15 0.14 0.12 -0.13 -0.12 -0.16 -0.12
GDP (USD)
Electric power consumption (kWh per capita)
Life expectancy at birth (years)
GDP per capita (USD)
Individuals using the Internet (% of population)
Birth rate, crude (per 1,000 people)
Infant mortality rate (per 1,000 live births)
Death rate, crude (per 1,000 people)
Unemployment (% of total labor force) (modeled ILO estimate)
Corr
1.0
0.5
0.0
-0.5
-1.0

General Overview

- **Color Coding**:
  - Positive correlations are shown in shades of blue.
  - Negative correlations are shown in shades of red.
  - The intensity of the color indicates the strength of the correlation.

Key Observations

1. **High Positive Correlations**:
   - **GDP per capita (USD) and Life expectancy at birth (years)**: Strong positive correlation (0.83), suggesting that higher GDP per capita is associated with higher life expectancy.
   - **GDP per capita (USD) and Individuals using the Internet (% of population)**: Strong positive correlation (0.87), indicating that wealthier countries tend to have higher internet usage rates.
   - **Life expectancy at birth (years) and Individuals using the Internet (% of population)**: Positive correlation (0.73), implying that increased internet usage is associated with longer life expectancy.
2. **Moderate Positive Correlations**:
   - **GDP per capita (USD) and Electric power consumption (kWh per capita)**: Positive correlation (0.72), suggesting that higher GDP per capita is linked with greater electric power consumption.
   - **Life expectancy at birth (years) and GDP (USD)**: Moderate correlation (0.58), indicating that countries with higher GDP tend to have longer life expectancy.
3. **Negative Correlations**:
   - **Infant mortality rate (per 1,000 live births) and Life expectancy at birth (years)**: Negative correlation (-0.75), showing that higher infant mortality rates are associated with lower life expectancy.
   - **Death rate, crude (per 1,000 people) and Life expectancy at birth (years)**: Moderate negative correlation (-0.51), suggesting that higher death rates are linked with lower life expectancy.
   - **Unemployment (% of total labor force) and Individuals using the Internet (% of population)**: Weak negative correlation (-0.14), indicating that higher unemployment rates are slightly associated with lower internet usage.

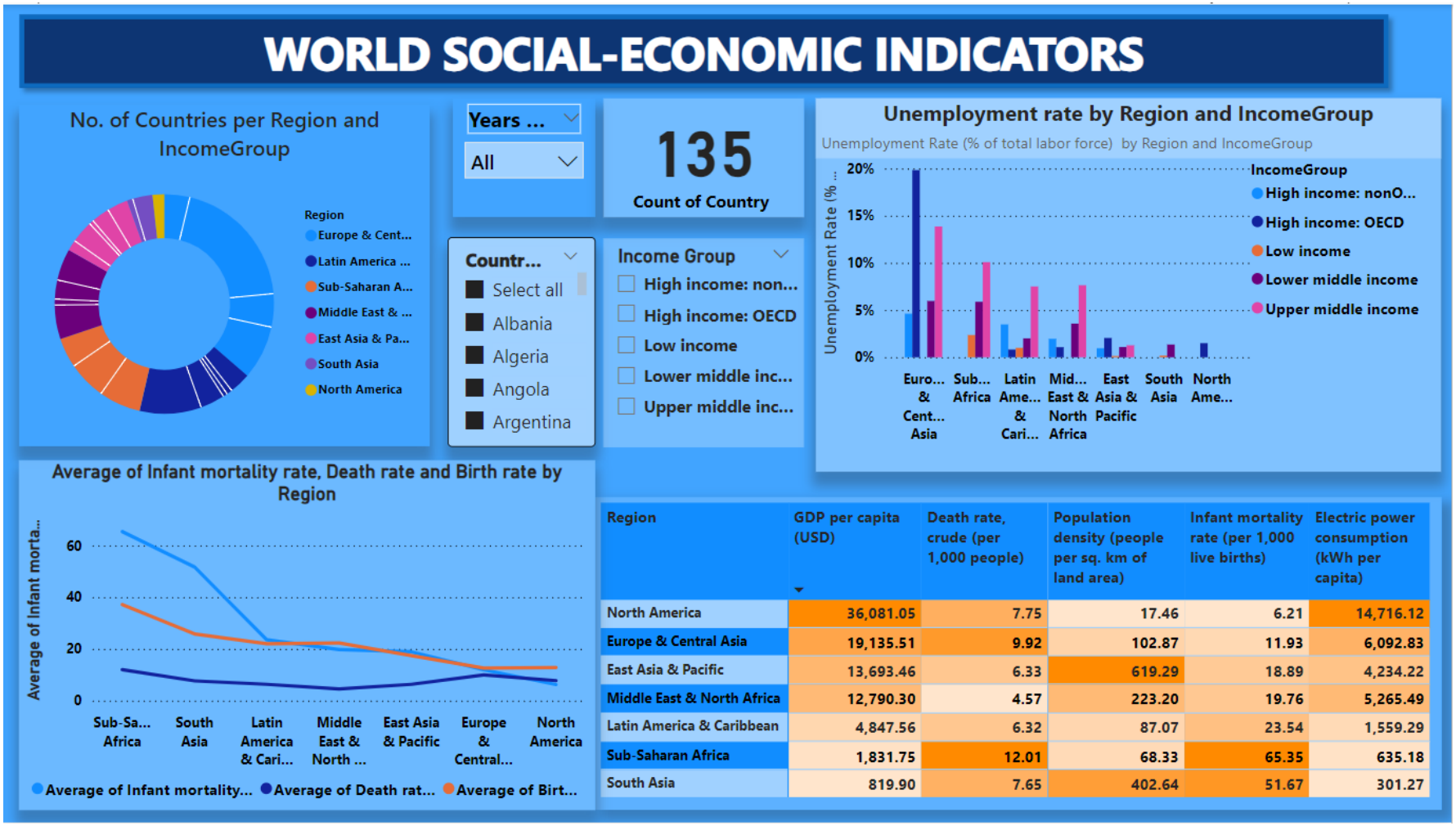
Additional Insights

- **Population density (people per sq. km of land area)**:
  - Has a weak negative correlation with GDP per capita (-0.10) and life expectancy (-0.12), suggesting that more densely populated countries might have slightly lower GDP per capita and life expectancy.
- **Electric power consumption and Individuals using the Internet**: Moderate positive correlation (0.58), indicating that higher electricity consumption is associated with increased internet usage.

Socioeconomic Implications

- **Economic Development**: Countries with higher GDP per capita tend to have better socio-economic indicators such as higher internet usage, longer life expectancy, and lower infant mortality rates.
- **Healthcare and Infrastructure**: Improved healthcare (evidenced by lower infant mortality rates and longer life expectancy) is associated with higher GDP per capita and better access to technology (internet usage).
- **Energy Consumption**: Economic prosperity (GDP per capita) is closely linked to higher energy consumption, which can be indicative of industrial and technological development.

Power Bi world social-economic indicators dashboard



Region	GDP per capita (USD)	Death rate, crude (per 1,000 people)	Population density (people per sq. km of land area)	Infant mortality rate (per 1,000 live births)	Electric power consumption (kWh per capita)
North America	36,081.05	7.75	17.46	6.21	14,716.12
Europe & Central Asia	19,135.51	9.92	102.87	11.93	6,092.83
East Asia & Pacific	13,693.46	6.33	619.29	18.89	4,234.22
Middle East & North Africa	12,790.30	4.57	223.20	19.76	5,265.49
Latin America & Caribbean	4,847.56	6.32	87.07	23.54	1,559.29
Sub-Saharan Africa	1,831.75	12.01	68.33	65.35	635.18
South Asia	819.90	7.65	402.64	51.67	301.27

Below is an analysis of the different elements of the dashboard

1. No. of Countries per Region and IncomeGroup

- **Visualization**: Donut chart showing the distribution of countries by region and income group.
- **Insight**: This visualization helps to understand the geographical and economic distribution of countries. For instance, it can reveal how many countries fall into each income category within a specific region, aiding in targeted economic analysis or policy-making.
- **Business Decision**: Businesses can use this information to identify potential markets for expansion based on income levels and regional distribution. For instance, focusing on high-income regions for premium products or low-income regions for affordable solutions.

2. Unemployment Rate by Region and IncomeGroup

- **Visualization**: Bar chart showing unemployment rates by region and income group.
- **Insight**:According to the chart, Europe & Central Asia have the highest unemployment rates, followed by regions like Sub-Saharan Africa and Latin America & Caribbean. This suggests that economic challenges might be more pronounced in these areas.
- **Business Decision**: High unemployment rates in Europe & Central Asia indicate economic instability which may affect market potential but also present opportunities for workforce availability. Businesses might consider investing in these regions to create job opportunities or tap into the available labor force. Alternatively, businesses should be cautious of economic instability in these areas when planning market entry.

3. Average of Infant Mortality Rate, Death Rate, and Birth Rate by Region

- **Visualization**: Line chart showing trends in infant mortality, death, and birth rates across regions.
- **Insight**: This visualization provides insight into the health and demographic trends in different regions. For instance, Sub-Saharan Africa shows higher infant mortality and death rates compared to other regions.

- **Business Decision**: Health-focused businesses and NGOs can use this data to identify regions in need of medical supplies, healthcare services, and support programs. Companies in the health sector might find opportunities for investment and expansion in regions with high infant mortality and death rates.

4. Socio-Economic Indicators Heatmap

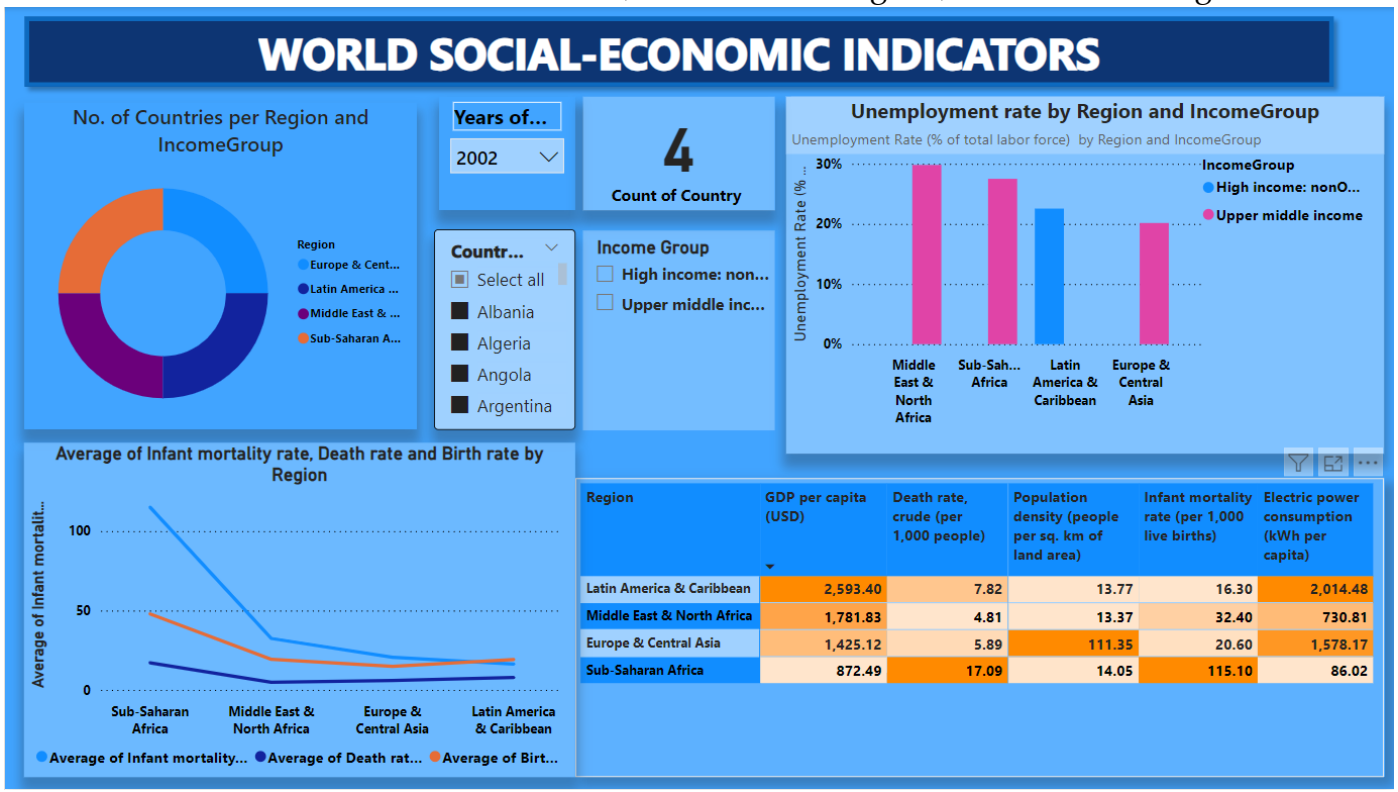
- **Visualization**: Heatmap displaying various socio-economic indicators such as GDP per capita, death rate, population density, infant mortality rate, and electric power consumption by region.
- **Insight**: This heatmap allows for a comparative analysis of socio-economic conditions across regions. The intensity of the colors helps to quickly identify regions with extreme values. For example, North America has the highest GDP per capita and electric power consumption, while Sub-Saharan Africa has the lowest GDP per capita and highest infant mortality rate.
- **Business Decision**: This data can inform various business strategies:
    - **Market Potential**: High GDP per capita regions (e.g., North America) can be targeted for high-end products and services.
    - **Infrastructure Development**: Regions with low electric power consumption (e.g., Sub-Saharan Africa) may offer opportunities for investment in infrastructure and renewable energy projects.
    - **Healthcare Initiatives**: Regions with high infant mortality rates (e.g., Sub-Saharan Africa) could benefit from healthcare investments and philanthropic activities.

Recommendations:

1. **Targeted Market Expansion**: Businesses should consider expanding into high-income regions for premium products and into low-income regions with affordable offerings tailored to local needs.
2. **Investment in Workforce Development**: Companies can leverage high unemployment rates in regions like Sub-Saharan Africa to establish operations, provided the economic and political conditions are stable.
3. **Healthcare and Social Initiatives**: Invest in regions with poor health indicators to improve community well-being and create a positive brand image. This can also open up new markets for healthcare products and services.
4. **Infrastructure Projects**: Consider opportunities in regions with low electric power consumption for infrastructure and energy projects, potentially in partnership with local governments or international organizations.

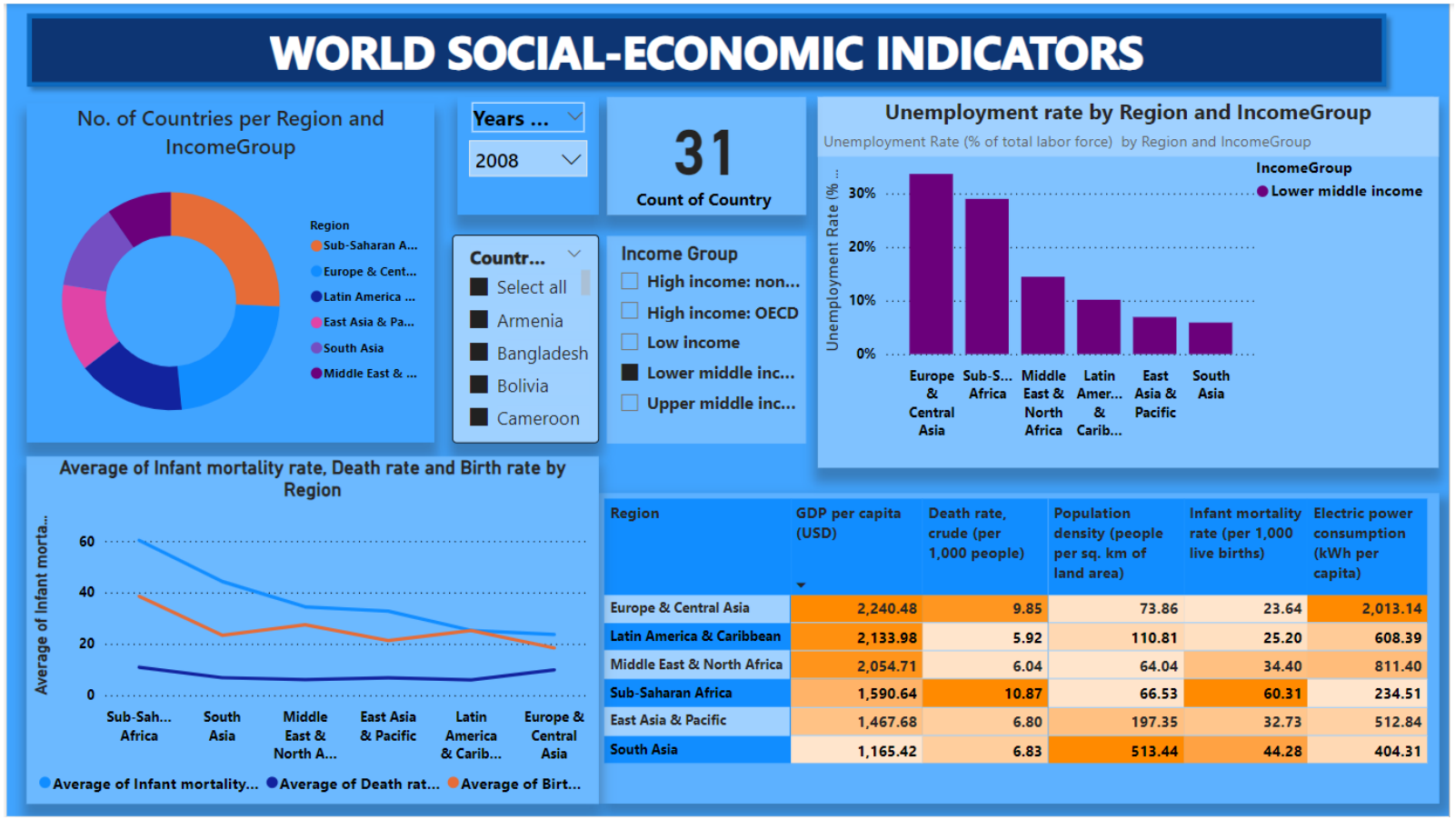
5. By analyzing these socio-economic indicators, businesses can make informed decisions about market entry, product development, and corporate social responsibility initiatives. The heatmap's visual intensity provides a quick reference to identify areas needing attention or presenting opportunities.

NOTE: A point to note is that the dashboard has 2 slicers ie one for years of interest and the other for country Name and a card to represent the number of countries that belong to a particular category selected. This enables the user to get statics for all countries(135 in total) and all the years in the dataset(1991 to 2014) or to trim down the data and get information for one country in a particular year or particular countries in particular years. For example this is how the dashboard will look if we want statistics for the year 2002 for Albania, Algeria, Angola and Argentina.



For the above snapshot the count is showing 4 because four countries have been selected. The income group slicer automatically changes to show the income groups the selected countries belong to for that particular year

The card count can further enable us to know how many countries fall in a particular income group in a given year or years as shown below.



The above snapshot is showing that in 2008 we had 31 countries belonging to the lower middle income group and the highest unemployment rates in that year was in countries from Europe and Central Asia followed by Sub-Saharan Africa. This and other kinds of visualisations can be generated from the dashboard depending on the needs of the user.